

# Sequence Preference in RNA Recognition by the Nucleoporin Nup153<sup>\*S</sup>

Received for publication, September 5, 2006, and in revised form, January 18, 2007. Published, JBC Papers in Press, January 22, 2007, DOI 10.1074/jbc.M608477200

Jennifer R. Ball<sup>1</sup>, Christian Dimaano<sup>1</sup>, Amber Bilak, Eydiejo Kurchan, M. Tracy Zundel, and Katharine S. Ullman<sup>2</sup>  
From the Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah 84112

The vertebrate nuclear pore protein Nup153 contains a novel RNA binding domain. This 150-amino acid region was previously found to bind preferentially to a panel of mRNAs when compared with structured RNAs, such as tRNA, U snRNA, and double-stranded RNA. The ability to broadly recognize mRNA led to the conclusion that the Nup153 RNA binding domain confers a general affinity for single-stranded RNA. Here, we have probed Nup153 RNA recognition to decipher how this unique RNA binding domain discriminates between potential targets. We first mapped the binding determinant within an RNA fragment that associates relatively robustly with the Nup153 RNA binding domain. We next designed synthetic RNA oligonucleotides to systematically delineate the features within this minimal RNA fragment that are key to Nup153 RNA-binding domain binding and demonstrated that the binding preferences of Nup153 do not reflect general preferences of an mRNA/single-stranded RNA-binding protein. We further found that the association between Nup153 and a cellular mRNA can be attributed to an interaction with specific subregions of the RNA. These results indicate that Nup153 can discriminate between mRNA and other classes of RNA transcripts due in part to direct recognition of a loose sequence motif. This information adds a new dimension to the interfaces that can contribute to recognition in mRNA export cargo selection and fate.

Nuclear pore complexes are macromolecular structures that bridge the inner and outer nuclear membranes to form a channel for nucleocytoplasmic traffic (1–3). Recent molecular characterization of pore complexes has revealed that they are comprised of only ~30 different proteins (4, 5), with multiples of eight copies of each protein forming the 8-fold symmetric structure characteristic of the nuclear pore complex. A small number of nucleoporins are restricted in localization to either the nuclear or cytoplasmic side of the pore, and their skewed

distribution contributes to distinct features on each of these faces: the nuclear basket structure and the cytoplasmic filaments. The observation that repetitive arrangement of a relatively limited number of proteins creates the elaborate pore structure suggests that each nucleoporin carries out multiple tasks, from providing structural scaffolding to contacting diverse cargo-receptor complexes as they transit through the pore.

The paradigm of multifunctional pore components is well illustrated by the nucleoporin Nup153. This pore protein plays roles in both import and export pathways (6–11). Consistent with this, Nup153 interacts with various transport receptors (7, 12–18), as well as with specific cargo (6, 10). Nup153 has also been found to be essential for the localization of other pore components (19) and is specifically thought to anchor the basket constituent TPR (20). Somewhat paradoxically, Nup153 has been shown to be dynamically localized to the nuclear pore (21, 22). This apparent contradiction, as well as the multiple roles implicated for Nup153, may be reconciled by the presence of distinct populations of this pore protein. Indeed, an analysis of Nup153 mobility indicated that a fraction of Nup153 is stably associated with the pore (Ref. 22, see also Ref. 21). To understand the various interactions that Nup153 can engage in to carry out its multiple roles, we previously probed this protein for RNA binding ability and consequently mapped a novel RNA binding domain to the unique N-terminal region of Nup153 (23). We further found that single-stranded RNA preferentially associates with this region (23, 24), suggesting that mRNA cargo en route to or through the nuclear pore may contact Nup153 both directly as well as through protein-protein contacts with mRNA-associated proteins.

In our previous study, we found that both full-length Nup153, as well as the Nup153 RBD<sup>3</sup> in isolation, bound non-discriminately to several mRNAs when tested in an *in vitro* binding assay (24). Heterogeneity in binding ability was observed only among mRNA fragments of 150 nucleotides or less. While this pointed toward an influence of length on recognition, we established that length alone was not the determining factor as structured RNAs, even if greater than 150 nucleotides, do not bind to the Nup153 RBD. To understand what targets an RNA for recognition by Nup153, and ultimately how and when this interaction occurs, we have probed the determinants of recognition within single-stranded RNA. We

\* This work was supported by National Institutes of Health Grant GM61275 (to K. S. U.) and a Ruth L. Kirschstein Individual National Service Award (to C. D.). Core facilities used in this study were partially supported by National Institutes of Health Grant P30 CA421014. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>S</sup> The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Tables S1 and S2.

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> To whom correspondence should be addressed: Dept. of Oncological Sciences, 2000 Circle of Hope, Huntsman Cancer Inst., University of Utah, Salt Lake City, UT 84112. Tel.: 801-585-7123; Fax: 801-585-0900; E-mail: [katharine.ullman@hci.utah.edu](mailto:katharine.ullman@hci.utah.edu).

<sup>3</sup> The abbreviations used are: RBD, RNA binding domain; GFP, green fluorescent protein; SELEX, systematic evolution of ligands by exponential enrichment; DHFR, dihydrofolate reductase; GST, glutathione S-transferase; hnRNP, heterogeneous nuclear ribonucleoprotein.

have found that Nup153 RBD binding is dependent on specific nucleotide distribution and arrangement. Although the requirements for binding are sequence based, the recognition motif requirements appear to be loose enough to be present in a large population of mRNA.

## EXPERIMENTAL PROCEDURES

**Transcription Templates and *in Vitro* Transcription**—For transcription of the RNA ladder, pBluescript-SK<sup>+</sup> (Stratagene) was used as a template to generate seven PCR products differing in length by 100 nucleotides each. All oligonucleotides are listed in supplemental Table S1. The forward primer, fLadder, hybridizes upstream of the T7 promoter and was combined with one of seven reverse primers (r100–r700) downstream of the T7 promoter. The actual length of these transcripts is two nucleotides greater due to two guanines encoded by the T7 promoter; for simplicity in nomenclature we did not include this in our numbering. pBluescript was also used to PCR amplify a template for transcription of 100-AU. Since the sequence was not adjacent to a transcriptional promoter, the T7 promoter was included in the forward primer (f100-AU). Templates were transcribed *in vitro* with the Maxiscript kit (Ambion) using T7 polymerase. RNA was radiolabeled with [ $\alpha$ -<sup>32</sup>P]UTP (800 Ci/mmol; PerkinElmer Life Sciences) during transcription. In the case of RNase H digestions, the 100- and 200-mer ladder transcripts were synthesized as above except MegaScript T7 polymerase (Ambion) was used to increase the RNA yield during transcription. DHFR fragments were generated similarly, using PCR to create a short template and MegaScript T7 polymerase to transcribe.

**Purification and 5' End Labeling of Synthetic RNA**—38 nucleotide long RNA sequences (see Fig. 4) were synthesized by The University of Utah DNA/RNA synthesis facility. The oligonucleotides were run on a 12% denaturing acrylamide gel. RNA was visualized with ethidium bromide and retrieved from gel slices by soaking crushed gel fragments in 0.5 M ammonium acetate, 1 mM EDTA overnight, followed by butanol extraction and ethanol precipitation. Each oligonucleotide was then 5' end-labeled following standard procedures.

**RNase H Digestion**—DNA oligonucleotides were designed to anneal to sites along the 100- and 200-mer ladder transcript. Each oligonucleotide was at least 12 nucleotides long and had a  $T_m \geq 40^\circ\text{C}$  (see supplemental Table S1). To cleave the RNA (see supplemental Table S2), half of a 20- $\mu\text{l}$  transcription reaction was annealed to RNase H oligonucleotide (30  $\mu\text{M}$ ) in 250 mM Tris, pH 7.5, 10 mM EDTA, 500 mM NaCl. Samples were heated at 68  $^\circ\text{C}$  for 10 min, followed by a 5  $^\circ\text{C}$  temperature drop every 1 min until the final temperature reached 30  $^\circ\text{C}$ . After incubation at 30  $^\circ\text{C}$  for 10 min, an equal volume of 2 $\times$  RNase H buffer (40 mM Tris, pH 7.5, 20 mM MgCl<sub>2</sub>, 100  $\mu\text{M}$  NaCl, 2  $\mu\text{M}$  dithiothreitol, 10 units of RNase H, 60  $\mu\text{g}/\text{ml}$  bovine serum albumin) was added to the annealing reaction. After the addition of RNase H, samples were incubated at 30  $^\circ\text{C}$  for 1 h. The reaction was stopped by the addition of 1  $\mu\text{l}$  0.5 M EDTA, pH 8.0 (25).

**Recombinant Protein Expression**—The constructs for His<sub>6</sub>-RBD and -GFP proteins, as well as their purification, was as described previously (23, 24). A construct encoding His<sub>6</sub>-GFP-

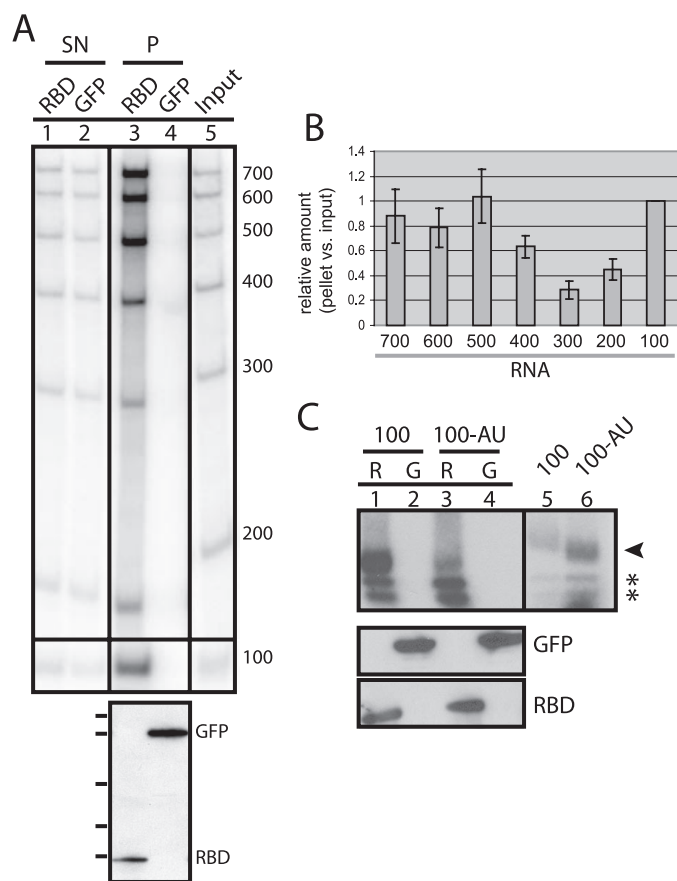
hnRNP A1 was a kind gift of Matthew Michael (Harvard University). A GST-RBD-6His construct was made using PCR to introduce a His<sub>6</sub> tag at the C-terminal end of a GST-RBD (amino acids 260–410 of *Xenopus* Nup153) fusion protein in a pGEX4T backbone vector. This protein was purified following standard native preparation procedures and imidazole elution. The His<sub>6</sub>-Nup153-NZ construct and its purification were described previously (construct 3 (26)).

**RNA Pulldown Procedure**—RNA pulldown assays were performed as described previously (24). Briefly, recombinant protein (~250 ng) was bound to anti-T7 antibody immobilized on agarose beads (Novagen) in binding buffer (100 mM KCl, 25 mM HEPES, 0.5% Triton X-100, 10 mg/ml bovine albumin). In some assays, this step was done the day before the RNA binding itself. In this case, the charged beads were stored overnight in binding buffer at 4  $^\circ\text{C}$ . Labeled RNA was added to the beads in the presence of binding buffer, 3 mg/ml heparin (Calbiochem) and 40 units of RNase Out (Invitrogen). After incubation for 5–60 min at room temperature, the beads were washed three times in wash buffer (200 mM KCl, 25 mM HEPES, 0.5% Triton X-100) and divided equally into two reactions for either RNA or protein analysis. Protein recovery was monitored by immunoblot with anti-T7 antibody. RNA was isolated from the remainder of the pellet and electrophoresed on 6 or 10% denaturing acrylamide gels.

**Gel Mobility Shift Assay**—End-labeled RNA was mixed with binding buffer containing 100 mM KCl, 25 mM HEPES, 0.5% Triton X-100, 0.5 mg/ml bovine albumin, 2 mg/ml heparin, 7.5% glycerol, 1 $\times$  Complete protease inhibitor (Calbiochem), and 120 units of RNase Out (Invitrogen). Recombinant proteins (~200 ng) were then added and incubated 1 h on ice. Samples were electrophoresed at 4  $^\circ\text{C}$  on a pre-run 6% native acrylamide gel ( $\frac{1}{2}\times$  TBE (0.045 M Tris borate, 1 mM EDTA), 2.5% glycerol, 37.5:1 acrylamide:bisacrylamide) for 2 h at 300 volts.

## RESULTS

**The RBD Associates with a Broad Range of RNA Sizes**—We previously observed a trend in which shorter RNAs (100–200 nucleotides) did not bind predictably to the Nup153 RBD as did longer RNAs (300–800 nucleotides) (24). To examine the influence of length more systematically, we generated a ladder of RNAs. Since we had found that the Nup153 RBD recognizes a wide range of single-stranded RNAs, we chose to derive the RNA ladder from plasmid sequence. A panel of seven transcripts was generated (100–700 nucleotides) and tested for binding to recombinant Nup153 RBD. Following immobilization of RBD and GFP proteins, the radioactively labeled RNAs were mixed and incubated with each protein in the presence of heparin sulfate. Samples were washed, and the beads were divided to analyze both protein recovery and bound RNA. The experiment was done in triplicate, with similar results in each case (a representative sample is shown in Fig. 1A, and all three are quantified in Fig. 1B). Recovery of GFP and RBD proteins was relatively equivalent, as assessed by immunoblot detection of a T7 epitope (Fig. 1A, lower panel). All RNAs, from 100 to 700 nucleotides, bound to recombinant RBD (Fig. 1A, lane 3) but not GFP (Fig. 1, lane 4). While the larger RNAs are greater in intensity due to more labeled nucleotide per transcript, when



**FIGURE 1. The Nup153 RNA binding domain binds to short RNAs with a nucleotide content biased toward GC.** *A*, recombinant T7-tagged RBD or GFP was immobilized with anti-T7 antibody conjugated to agarose beads. These beads were then incubated with a ladder of radiolabeled RNAs, and following washing steps, the beads were divided in two and analyzed for protein and RNA recovery. 50% of the pellet material was electrophoresed on an SDS-polyacrylamide gel and subjected to immunoblot detection with anti-T7 antibody (lower panel). RNA was isolated from the other half of the pellet and run on a 6% denaturing polyacrylamide gel (upper panel, lanes 3 and 4), along with 2% of each binding reaction supernatant (lanes 1 and 2) and 2% of the input material (lane 5). *B*, RNA recovery in three separate reactions was quantitated by phosphorimager analysis and the relative binding of each fragment is graphed. For each pellet, the ratio of the fragments in the pellet was calculated relative to the 100-mer. The same calculation was made for the input. The ratios in the pellet were then compared with input ratios. *C*, either the 100-mer ladder transcript or a 100-mer that was chosen for its enrichment in AU (100-AU) content was tested for binding as described for *A*. RBD (*R*) and GFP (*G*) pellets are shown in lanes 1–4, with RNA recovery in the upper panel and protein recovery in the lower panels. The RNA input (2%) is in lanes 5 and 6. The asterisk indicates breakdown products from DHFR RNA (data not shown) that had been included in the reaction.

binding is considered quantitatively, the recovery of each individual RNA does not correlate with its length (Fig. 1*B*). The relatively poor recovery of the 200- and 300-mer suggests that certain sequences may have a negative effect on Nup153 RBD binding. Most importantly with regard to the question at hand, these results indicated that there is not an intrinsic cutoff to recognition at 100–200 nucleotides.

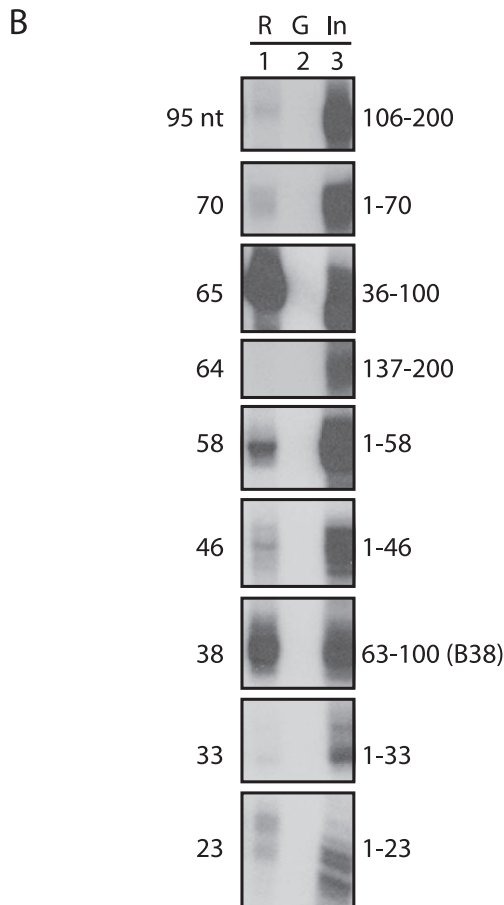
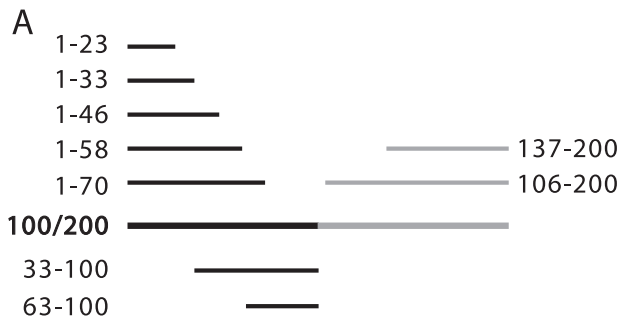
**The RBD Associates with GC-rich RNA**—The observation that the RBD can bind the 100-mer ladder transcript gave us an opportunity to dissect binding requirements. When we examined this particular RNA for salient features that might be important to recognition, we noted a bias in guanosine and cytosine content (19% A, 20% U, 30% C, 31% G). To test whether

the nucleotide distribution is important for recognition by the Nup153 RBD, we generated an independent 100-mer transcript that had an AU versus a GC bias. When the 100-AU (26% A, 32% U, 22% C, 20% G) was tested alongside the 100-mer ladder transcript, there was a clear difference in RBD binding ability. Again, GFP and RBD recombinant proteins were immobilized on agarose beads and then incubated with transcripts. Following the washing steps, RNA that remained bound was analyzed. The 100-mer ladder transcript associated with the RBD (Fig. 1*C*, lane 1) more robustly than did 100-AU (Fig. 1*C*, lane 3), especially considering there was more 100-AU in the input (lane 6) than 100-mer ladder transcript (lane 5). This demonstrates that the RBD is indeed selective in its recognition and shows a binding preference for RNA targets enriched in guanosine and cytosine.

**The RBD of Nup153 Associates with Defined Regions within 100-mer Ladder Transcript**—To further dissect the determinants of RBD association, we fragmented the 100-mer and 200-mer ladder transcripts with oligonucleotide-directed RNase H digestion. These defined regions were then incubated with immobilized RBD and GFP to further delineate the recognition basis for preferred RNA substrates. A wide range of fragment sizes was generated in separate reactions (see supplemental Table S2 and Fig. 2*A*) and are depicted together in a panel (Fig. 2*B*). These fragments do not always run as tight bands, likely due to some imprecision in the exact nucleotides digested by RNase H at each end. Comparison of binding among these fragments definitively uncoupled length from RNA binding ability (Fig. 2*B*, lane 1). Certain sequences, even though longer, lacked binding activity (for example, Fig. 2*B*, 106–200). Rather, the ability to associate with the RBD clearly mapped to a particular region within the 100-mer ladder transcript, with a 38-mer corresponding to the 3' end of this 100-mer being the shortest fragment with robust recognition (Fig. 2, lane 1, 63–100 fragment, for simplicity we will refer to this minimal fragment as B38, for binding 38-mer).

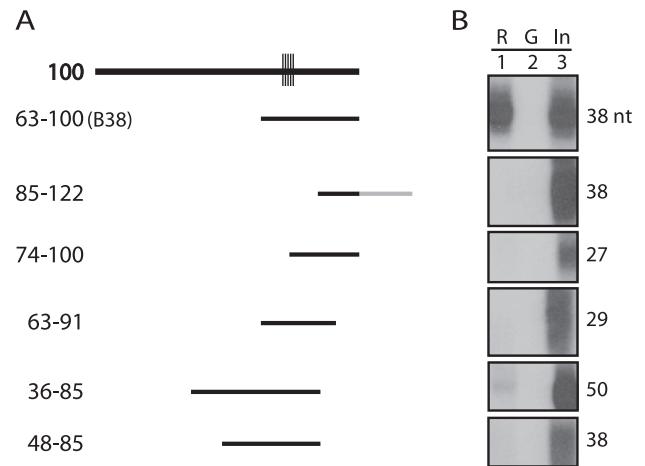
This analysis did not support the idea that nucleotide distribution alone is sufficient to create a RBD recognition motif. Specifically, although B38, the 38-mer mentioned above, did have a bias toward GC content (16% A, 16% U, 34% C, 34% G), fragments that did not bind well had a similar profile. For example, 1–46 (Fig. 2) is 19% A, 17% U, 29% C, 35% G, and the 85–122 (Fig. 3) fragment is 13% A, 21% U, 34% C, 32% G. Thus, Nup153 RBD recognition cannot be accounted for by either length or absolute GC content. Therefore, despite the general nature of this interaction, there appears to be a sequence component that contributes to RNA recognition by Nup153.

We noted that a stretch of five guanines was present in the minimal binding fragment (B38, see Table 1 and indicated by hatch marks in Fig. 3). Given that Nup153 had been previously found to bind to poly(G) (8, 23), this clustering was a candidate for being a core motif in RBD recognition. Indeed, when a 38-mer with a similar GC content but lacking the G cluster was tested, it did not have binding activity (Fig. 3*B*, lane 1, 85–122). Similarly, when the minimal 38-mer was truncated at the 5' end such that this G-cluster was absent, binding ability was significantly decreased (Fig. 3*B*, lane 1, 74–100). However, we also truncated the minimal 38-mer from the 3' end by nine nucleo-



**FIGURE 2. Mapping the interaction between the Nup153 RBD and the 100-mer ladder transcript.** *A*, short sequences generated by oligonucleotide-directed RNase H cleavage of the 100- and 200-mer ladder fragments are schematically depicted (see supplemental Table S2 for details). *B*, binding of these particular sequences was tested as described for Fig. 1*A*, and individual results are combined in these panels, with RBD pellets (50%) in lane 1, GFP pellets (50%) in lane 2, and input (*In*) (2%) in lane 3. The nucleotide length of each RNA is indicated at the left of each panel, and nucleotide coordinates within the 200 nucleotide region are indicated at the right of each panel. The 38-mer, which is the minimal binding fragment, is indicated as B38.

tides and found that, although the G-cluster was intact in this fragment, it was not sufficient for binding (Fig. 3*B*, lane 1, 63–91). Since we could not distinguish whether the decrease in binding in the latter two cases was due to falling below a length threshold or missing particular sequence determinants, we tested other fragments that were 38 nucleotides or longer and contained the G-cluster but not the entire minimal 38-mer. These, too, did not bind well to the RBD (Fig. 3*B*, lane 1, 36–85 and 48–85). These binding patterns indicate that although G-clustering may contribute to RNA recognition by the



**FIGURE 3. Thirty-eight nucleotides at the 3' end of the 100-mer ladder is the minimal recognition motif.** *A*, short sequences derived by oligonucleotide-directed RNase H cleavage are depicted schematically. 85–122 and 63–91 were generated from the 200-mer ladder transcript (see supplemental Table S2). A cluster of five guanines is indicated by hatch marks. *B*, binding of these particular sequences was tested as in Fig. 1*A*, and individual results are combined in these panels, with RBD pellets (50%) in lane 1, GFP pellets (50%) in lane 2, and input (*In*) (2%) in lane 3. Each RNA corresponds to the fragment depicted to the left (*A*), and the nucleotide length of each RNA is indicated at the right of each panel.

Nup153 RBD, it is not sufficient for robust binding, and therefore, additional features within the minimal B38 binding fragment are likely key to recognition. Retention of Nup153 on a poly(G) matrix may be due to the presence of multiple low affinity sites. Consistent with this, we found that a 31 nucleotide RNA oligonucleotide with two stretches of five guanines could bind to the RBD above background but very weakly compared with the B38 RNA (data not shown).

**Sequence Requirements in Nup153 Recognition**—To assess how the RNA binding domain within Nup153 preferentially associates with B38, we designed a series of synthetic RNA oligos and subjected these to a binding analysis. In each case, the oligonucleotides were gel-purified, end-labeled, and incubated with immobilized RBD. Consistent with the information derived from 100-mer ladder fragments, we found that if the distribution of guanosine within this 38-mer, including the 5G cluster, was maintained within an altered backbone, the resulting RNA was not recovered with the RBD in a pull-down assay (Fig. 4, 1*G5*, lane 3). Along the same lines, distributing guanines into clusters of 2 or 3 also resulted in little to no binding to the Nup153 RBD in this assay (Fig. 4, 2–3*G*, lane 3). Likewise, a simple repetitive sequence of AUCG was unable to associate with the RBD (Fig. 4, AUCG, lane 3). In contrast, a 38-mer in which the cytidine-guanosine clustering matched the pattern with B38 bound with equivalent efficiency as B38 (Fig. 4, G5GC, lane 3). The major difference between G5GC, which has binding activity, and 1*G5*, which does not have binding activity, is the distribution of cytidine content within a similar framework anchored by G-rich regions. Thus, Nup153 RNA recognition appears to involve motifs that are particular arrangements of both guanosine and cytidine.

**Comparison with Another mRNA-binding Protein, hnRNP A1**—Two important questions raised by the results of this binding analysis are: 1) do the RNAs that bind preferentially to the

TABLE 1

RNA fragments tested for binding to the Nup153 RBD

RNA with Nup153-RBD binding activity is indicated in bold.

RNA	Sequence
B38	<b>UGCAGCCCGGGGAUCCACUAGUUAAGAGCGGCCGCC</b>
G5GC	<b>AUAUACCCGGGGGAUAUAUAUAUAUGAGCGGCCGCC</b>
1G5	AGUCGAUCGGGGGAUCAUCAUGCAUCAGUGCGGAUGCA
2-3G	AUCAUCAUGGGCAGGUCAUCAGGGUCGGAUCAGGGUCU
AUCG	GAUGCAGUCGAUGCAGUCGAUGCAGUCGAUGCAGUCGA
A1	UGUGAUAGGGACUUAGGGUG
DHFR-1	GGAUGGUUCGACCAUUGAACUGCAUCGUCGCGGUGUCCAAAAUAUGGGGAUUGGCAAGAACG
DHFR-2	GGGGGAGACCUACCCUGGCCUCCGCUCAGGAACGAGUUAAGUACUCCAAAGAAUGACCACAACC
DHFR-3	<b>GGUGAAGGUAACAGAAUCUGGUGAUUAUGGGUAGGAAAACCUUGGUUCUC</b>
DHFR-4	GGCCAUUCUGAGAAGAAUCGACCUUUAAGGACAGAAUUAUAUGUUCUCAG

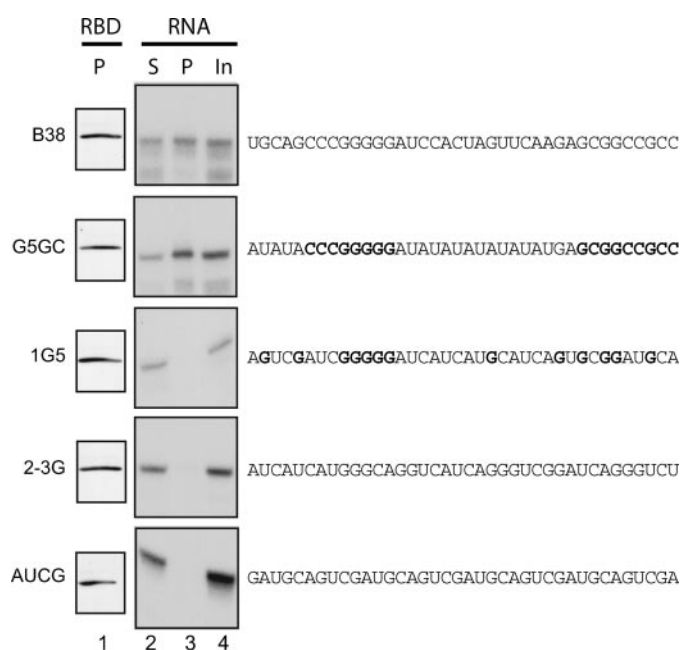


FIGURE 4. Two regions of GC content within B38 are key to recognition by the Nup153 RBD. RNA oligonucleotides were synthesized, gel-purified, radiolabeled, and subjected to binding analysis as described for Fig. 1A. Relative RBD recovery, as assessed by anti-T7 immunoblot, is shown in the left panels (lane 1). RNA was isolated from 4% of the supernatant (lane 2), 50% of the pellet (lane 3), and 2% of the input (lane 4) and run on a 10% denaturing acrylamide gel. Sequences are indicated at the right of each gel. Nucleotides with positions purposefully conserved with B38 are in bold.

Nup153 RBD reflect RNA targets that are generally recognized by ssRNA-binding proteins? and 2) how does the binding ability of Nup153 compare with well characterized mRNA-associated proteins? To address these two questions, we compared RNA binding by hnRNP A1 to the Nup153 RBD. hnRNP A1 associates with mRNA from an early point in its biogenesis (27) and has well characterized RNA binding ability, which is attributed to two tandem RNA recognition motifs. An electrophoretic mobility shift assay was used to monitor association between hnRNP A1 and various RNAs. The Nup153 binder, B38 (Figs. 2–4 and 5, lane 3), was similarly able to associate with hnRNP A1 (Fig. 5, lane 2). The specificity of the RNA-protein interaction is underscored by the lack of gel shift with either a recombinant protein derived from an adjacent region of Nup153 (designated NZ and containing amino acids 436–717 of xNup153; Fig. 5, lane 4) or with GST itself (lane 5).

Despite overlap in RNA recognition by Nup153 RBD and hnRNP A1, clear distinctions between the two were also apparent. For instance, the oligonucleotide G5GC, which preserves

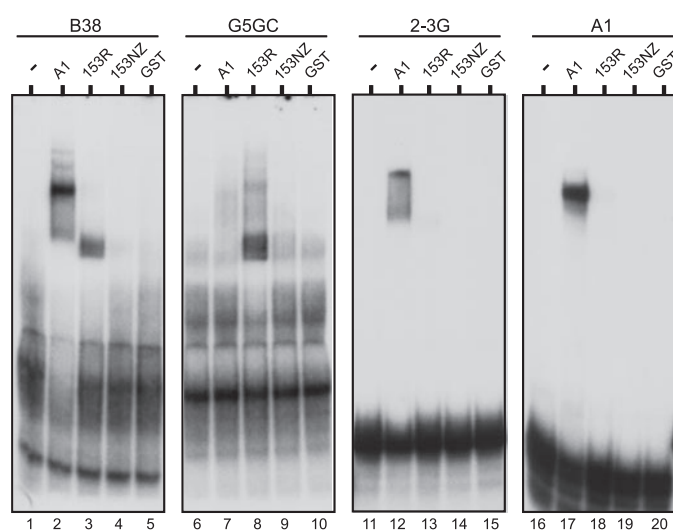
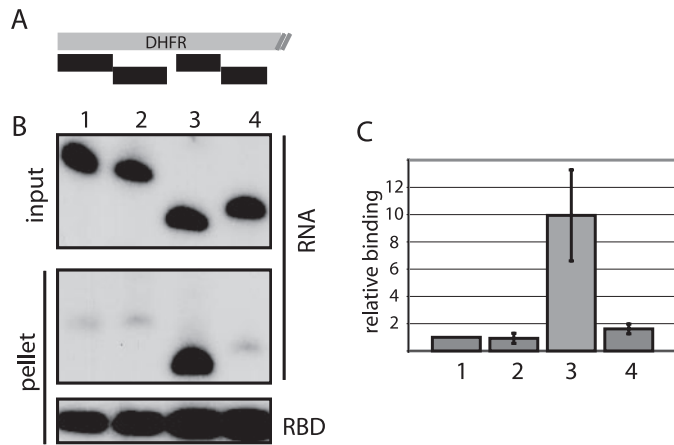


FIGURE 5. Nup153 RBD recognition preferences are distinct from those of hnRNP A1. Gel shift analysis was performed with a panel of end-labeled oligonucleotides. B38 (lanes 1–5), G5GC (lanes 6–10), and 2–3G (lanes 11–15) are described in the legend to Fig. 4. The A1 oligonucleotide (lanes 16–20) is a sequence previously determined to be efficiently recognized by hnRNP A1. RNA was incubated without protein (lanes 1, 6, 11, and 16) or with either His<sub>6</sub>-hnRNP A1 (lanes 2, 7, 12, and 17), GST-153RBD-His<sub>6</sub> (lanes 3, 8, 13, and 18), His<sub>6</sub>-153NZ (lanes 4, 9, 14, and 19), or GST (lanes 5, 10, 15, and 20).

the GC distribution of B38 in an altered backbone and remains able to associate with the Nup153 RBD, was not recognized by hnRNP A1 (Fig. 5, lane 7). And 2–3G, in which the guanosine content of B38 has been spread throughout the 38-mer in short clusters, is targeted by hnRNP A1, although it has lost the ability to interact with Nup153. Likewise, an optimal hnRNP A1 target, previously identified in SELEX experiments (28), does not interact with the Nup153 RBD (Fig. 5, lane 18), although it interacts as expected with hnRNP A1 (lane 17). Thus, the Nup153 RBD has a preference for particular sequences and its binding activity is distinctive from another ssRNA binding protein. Furthermore, this comparative gel shift analysis indicates that the RBD found within the nucleoporin Nup153 interacts with RNA within the same range of affinity as hnRNP A1.

**Recognition of Cellular RNA Targets**—Our results suggest that although the Nup153 RBD can interact with several mRNAs tested in an *in vitro* binding assay (24), this apparent general affinity for mRNA is likely mediated by specific interactions with particular sequences that are represented within mRNA. To test this hypothesis, we analyzed the interaction between the Nup153 RBD and DHFR mRNA, an RNA previously shown to be recognized by Nup153 (24). To determine whether the Nup153 RBD associates with particular sequences



**FIGURE 6. The Nup153 RBD associates with a particular region of DHFR mRNA.** *A*, the first 230 nucleotides of DHFR mRNA were divided into four fragments (labeled 1–4), each individually transcribed *in vitro*. See Table 1 for sequences. *B*, a binding assay was performed as described in the legend to Fig. 1. Input RNAs are shown in the upper panel, and RNA that remained associated with the RBD is shown in the middle panel. Immunoblot analysis of the RBD itself is shown in the lower panel. *C*, results of three independent experiments are shown graphically; binding is quantitated relative to the DHFR-1 fragment.

within DHFR mRNA, we took the proximal third of the DHFR open reading frame and subdivided it into four fragments of about 60 nucleotides each (Table 1). These RNA fragments were then tested for their ability to associate with the Nup153 RBD in a pull-down assay. Only one of the four subregions of DHFR RNA interacted robustly with the Nup153 RBD (Fig. 6, lane 3, middle panel). This result supports the notion that direct contact between Nup153 and mRNA is dictated by recognition of motifs that are present at some frequency within an mRNA. It is important to note, however, that although there is a sequence component to this interaction, the evidence indicates that such motifs are present relatively frequently in RNA. Specifically, the Nup153 RBD was previously found to associate with eight mRNAs that were randomly selected (24), and in the present analysis, we found that Nup153 can interact additionally with regions of DHFR outside of the binding sequence identified in the first 230 nucleotides (data not shown). Thus, there are likely to be relatively loose requirements in what constitutes a Nup153 recognition motif. Consistent with this, there is no clear alignment between the DHFR fragment that binds the Nup153 RBD and the synthetic oligonucleotides that were used to delineate parameters of recognition (see Table 1).

## DISCUSSION

In our previous characterization of the Nup153 RNA binding domain, we observed that this pore protein associates preferentially with a population of RNA, namely mRNA, that is predicted to be relatively unstructured (24). Here, we have determined that the molecular basis for recognition of RNA is in fact a sequence preference. Our results point toward there being only loose constraints on motifs recognized by Nup153, consistent with the observation that many mRNAs can be recognized by Nup153. Absolute affinity as well as the specific context of these sites is likely to have important influences that remain to be elucidated. Indeed, although binding sequences may be prevalent within RNA, even other than mRNA, recog-

nition by Nup153 is likely to be constrained by secondary structure of RNA, which we have found is generally a poor target for recognition (24). The exact nature of motifs recognized by Nup153 have yet to be fully elucidated (see below), and it remains formally possible that the sequence biases which we have observed reflect a particular local secondary structure that is in fact recognized by Nup153. Putting various binding and non-binding sequences through structural prediction programs, however, does not provide any correlation with a structure or folding energy, leading us to postulate that the primary sequence, and its accessibility based on context, is the feature that directs contact between Nup153 and RNA.

This new information about a novel RNA binding domain within the vertebrate pore protein Nup153 will be an important consideration when deciphering how RNA cargo moves to and through the nuclear pore complex. It has been appreciated for several years that different classes of RNAs have different requirements for nucleocytoplasmic trafficking (29, 30). This has largely been attributed to different sets of soluble factors recognizing and ferrying RNA cargo through the pore. Our new results suggest that particular regions within mRNA may make direct contact with the pore protein Nup153, in conjunction with contacts mediated by mRNA adaptor/receptor proteins. Such a multifaceted interface would further distinguish mRNA cargo from other classes of RNA cargo.

The search for an “mRNA identity element,” in fact, has led to the conclusion that unstructured RNA over a particular length threshold, roughly defined around 300 nucleotides, is sufficient to direct an RNA to an mRNA export path (31–34). The nature of this length requirement has not been addressed as yet but may reflect, at least in part, a requirement for sequence that mediates contact with Nup153. Alternatively, rather than being a core element of mRNA export, contact between motifs within mRNA and Nup153 could aid in discriminating among cargo and influence downstream mRNA fates. With information in hand from this current study, it should now be possible to test these and other models.

The results presented here suggest that approaches such as SELEX will be promising avenues to pursue, both to gain a better understanding of the range of possible Nup153 recognition motifs and to identify short RNAs with superphysiological affinity. Such high affinity sequences could then be introduced into cells to specifically block the interface between Nup153 and endogenous RNAs, allowing a functional dissection of how this contact impacts the fate of mRNA. A large collection of SELEX-generated sequence information may also make it possible to take a bioinformatics approach to classifying cellular transcripts for their ability to bind Nup153 and to correlate this information with functional differences in RNA subsets.

An emerging theme in nuclear organization is the physical proximity of pore proteins and active chromosomal regions (35–39). These intimate connections, as well as other well documented links between various steps in RNA biogenesis (40, 41), suggest that pore proteins may be involved from an early point in RNA production. Thus, rather than envisioning a fully packaged mRNP encountering pore proteins only at the time of translocation through the nuclear pore complex, we must consider the role of pore proteins during transcription, processing,

and transit to the pore (42). This is further underscored in the case of Nup153 by the transcription dependence of its dynamics at the pore (22). The finding reported here that Nup153 has an inherent ability to distinguish among RNAs creates a new inroad into elucidating the role of the pore and how individual pore components contribute selectivity to pore function.

*Acknowledgments*—We thank Jason Woodbury for expert technical assistance with protein purifications and Amy Prunuske for critical reading of the manuscript. We also thank Dr. Nicholas Davidson for providing laboratory space to Jennifer R. Ball.

### REFERENCES

- Tran, E. J., and Wentle, S. R. (2006) *Cell* **125**, 1041–1053
- Weis, K. (2003) *Cell* **112**, 441–451
- Lim, R. Y., and Fahrenkrog, B. (2006) *Curr. Opin. Cell Biol.* **18**, 342–347
- Rout, M. P., Aitchison, J. D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B. T. (2000) *J. Cell Biol.* **148**, 635–651
- Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., Chait, B. T., and Matunis, M. J. (2002) *J. Cell Biol.* **158**, 915–927
- Marg, A., Shan, Y., Meyer, T., Meissner, T., Brandenburg, M., and Vinkemeier, U. (2004) *J. Cell Biol.* **165**, 823–833
- Shah, S., and Forbes, D. J. (1998) *Curr. Biol.* **8**, 1376–1386
- Ullman, K. S., Shah, S., Powers, M. A., and Forbes, D. J. (1999) *Mol. Biol. Cell* **10**, 649–664
- Bastos, R., Lin, A., Enarson, M., and Burke, B. (1996) *J. Cell Biol.* **134**, 1141–1156
- Xu, L., Kang, Y., Col, S., and Massague, J. (2002) *Mol. Cell* **10**, 271–282
- Ball, J. R., and Ullman, K. S. (2005) *Chromosoma (Berl.)* **114**, 319–330
- Nakielnny, S., Shaikh, S., Burke, B., and Dreyfuss, G. (1999) *EMBO J.* **18**, 1982–1995
- Moroianu, J., Blobel, G., and Radu, A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 9699–9704
- Bachi, A., Braun, I. C., Rodrigues, J. P., Pante, N., Ribbeck, K., von Kobbe, C., Kutay, U., Wilm, M., Gorlich, D., Carmo-Fonseca, M., and Izaurralde, E. (2000) *RNA (N. Y.)* **6**, 136–158
- Ben-Efraim, I., and Gerace, L. (2001) *J. Cell Biol.* **152**, 411–418
- Schmitt, I., and Gerace, L. (2001) *J. Biol. Chem.* **276**, 42355–42363
- Smitherman, M., Lee, K., Swanger, J., Kapur, R., and Clurman, B. E. (2000) *Mol. Cell Biol.* **20**, 5631–5642
- Yaseen, N. R., and Blobel, G. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 4451–4456
- Walther, T. C., Fornerod, M., Pickersgill, H., Goldberg, M., Allen, T. D., and Mattaj, I. W. (2001) *EMBO J.* **20**, 5703–5714
- Hase, M. E., and Cordes, V. C. (2003) *Mol. Biol. Cell* **14**, 1923–1940
- Daigle, N., Beaudouin, J., Hartnell, L., Imreh, G., Hallberg, E., Lippincott-Schwartz, J., and Ellenberg, J. (2001) *J. Cell Biol.* **154**, 71–84
- Griffis, E. R., Craige, B., Dimaano, C., Ullman, K. S., and Powers, M. A. (2004) *Mol. Biol. Cell* **15**, 1991–2002
- Dimaano, C., Ball, J. R., Prunuske, A. J., and Ullman, K. S. (2001) *J. Biol. Chem.* **276**, 45349–45357
- Ball, J. R., Dimaano, C., and Ullman, K. S. (2004) *RNA (N. Y.)* **10**, 19–27
- Muhlrad, D., and Parker, R. (1992) *Genes Dev.* **6**, 2100–2111
- Shah, S., Tugendreich, S., and Forbes, D. (1998) *J. Cell Biol.* **141**, 31–49
- Visa, N., Alzhanova-Ericsson, A. T., Sun, X., Kiseleva, E., Bjorkroth, B., Wurtz, T., and Daneholt, B. (1996) *Cell* **84**, 253–264
- Burd, C. G., and Dreyfuss, G. (1994) *EMBO J.* **13**, 1197–1204
- Jarmolowski, A., Boelens, W. C., Izaurralde, E., and Mattaj, I. W. (1994) *J. Cell Biol.* **124**, 627–635
- Pokrywka, N. J., and Goldfarb, D. S. (1995) *J. Biol. Chem.* **270**, 3619–3624
- Ohno, M., Segref, A., Kuersten, S., and Mattaj, I. W. (2002) *Mol. Cell* **9**, 659–671
- Ullman, K. S. (2002) *Curr. Biol.* **12**, R461–R463
- Masuyama, K., Taniguchi, I., Kataoka, N., and Ohno, M. (2004) *Genes Dev.* **18**, 2074–2085
- Rodrigues, J. P., Rode, M., Gatfield, D., Blencowe, B. J., Carmo-Fonseca, M., and Izaurralde, E. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1030–1035
- Galy, V., Olivo-Marin, J. C., Scherthan, H., Doye, V., Rascalou, N., and Nehrbass, U. (2000) *Nature* **403**, 108–112
- Casolari, J. M., Brown, C. R., Komili, S., West, J., Hieronymus, H., and Silver, P. A. (2004) *Cell* **117**, 427–439
- Cabal, G. G., Genovesio, A., Rodriguez-Navarro, S., Zimmer, C., Gadal, O., Lesne, A., Buc, H., Feuerbach-Fournier, F., Olivo-Marin, J. C., Hurt, E. C., and Nehrbass, U. (2006) *Nature* **441**, 770–773
- Menon, B. B., Sarma, N. J., Pasula, S., Deminoff, S. J., Willis, K. A., Barbara, K. E., Andrews, B., and Santangelo, G. M. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5749–5754
- Taddei, A., Van Houwe, G., Hediger, F., Kalck, V., Cubizolles, F., Schober, H., and Gasser, S. M. (2006) *Nature* **441**, 774–778
- Cole, C. N., and Scarcelli, J. J. (2006) *Curr. Opin. Cell Biol.* **18**, 299–306
- Vinciguerra, P., and Stutz, F. (2004) *Curr. Opin. Cell Biol.* **16**, 285–292
- Dimaano, C., and Ullman, K. S. (2004) *Mol. Cell Biol.* **24**, 3069–3076